

УДК 543.51:621.384.8

ПРОГРАММА ОБРАБОТКИ МАСС-СПЕКТРОВ

А.А.Соломеин, Б.А.Калинин, Л.М.Александров
Уральский государственный технический университет–УПИ
620002, Екатеринбург, Мира, 19
alexey@umc.ustu.ru

Поступила в редакцию 2 декабря 2002 г.

Разработана программа для обработки масс-спектра, представления информации по параметрам пиков, калибровки шкалы масс и для отображения дефектов массы углеводородов и фторуглеродов. Рассмотрен алгоритм обработки и интерфейс программы.

Соломеин Алексей Александрович – аспирант 1-го года кафедры мультимедиа технологий радиотехнического факультета УГТУ-УПИ.

Область научных интересов: масс-спектрометрия, программирование, мультимедиа технологии, оптимизация вычислений.

Калинин Борис Алексеевич – доцент кафедры молекулярной физики физико-технического факультета УГТУ-УПИ, кандидат физико-математических наук.

Область научных интересов: масс-спектрометрия.

Автор 30 статей.

Александров Леонид Михайлович – студент 4-го курса кафедры молекулярной физики физико-технического факультета УГТУ-УПИ.

Область научных интересов: масс-спектрометрия.

Разработанная программа написана в инструментальной среде Borland Delphi 5 под операционную систему Windows и предназначена для обработки спектра и вывода информации по параметрам пиков, калибровки шкалы масс и для отображения дефектов массы углеводородов и фторуглеродов. Программа имеет удобный интерфейс и отображает исходный и обработанный масс-спектр (рис. 1).

Масс-спектр поступает на обработку в виде текстового файла из двух столбцов: первый – масса в атомных единицах массы (а.е.м.) и второй – интенсивность ионного тока, ион/с.

Суть обработки состоит в разбиении всего масс-спектра на зоны пиков и обработки каждого пика путем аппроксимации его функцией Гаусса по методу наименьших квадратов (МНК). МНК реализован по способу определителей [1, с.258] с последовательным приближением параметров к минимуму суммы квадратов отклонений. Функция Гаусса выбрана в следующем виде:

$$f(M) = A + B \cdot e^{-\left(1.6651 \cdot R \cdot \left(\frac{M_0}{M} - 1\right)\right)^2}, \quad (1)$$

где f – интенсивность ионного тока, ион/с; M – массовая координата спектра, а.е.м.; A – фон в месте расположения пика, ион/с; B – высота пика, отсчитанная от уровня фона, ион/с; M_0 – положение центра пика, а.е.м.; R – безразмерный параметр, определяющий форму пика и равный отношению M_0 к ширине пика на половине высоты.

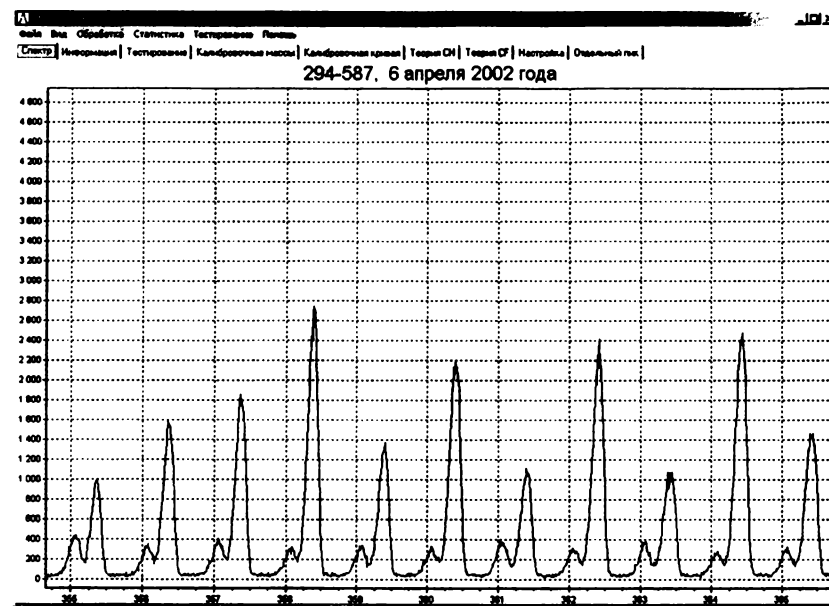


Рис.1. Вкладка "Спектр"

Один пик (~100 точек) аппроксимируется функцией Гаусса в среднем в течение 30 мс (на ПК с частотой 700 МГц). Зона пика имеет левую границу, максимальное значение (после многократных усреднений оно уменьшается) и правую границу (отмечены точками на рис. 2).

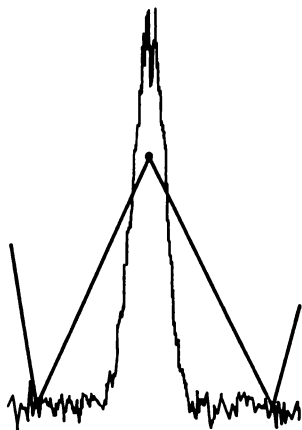


Рис.2. Пик и его зона

Определение зон пиков выполняется в две стадии (последовательность которых названа предварительной обработкой): многократное усреднение спектра и замена линиями возрастающих и убывающих участков. Последующая (последняя) обработка состоит из трех стадий: отсеивание зон пиков, чьи высоты меньше первого непопавшего пика (выбираемого пользователем), замена линиями зон помеховых пиков и отдельная обработка каждого пика в пределах его зоны.

В программе существует 9 вкладок: "Спектр", "Информация", "Тестирование", "Калибровочные

массы", "Калибровочная кривая", "Теория СН", "Теория CF", "Настройка" и "Отдельный пик".

На вкладке "Спектр" (Рис. 1) отображается график масс-спектра. Имеется возможность передвижения и увеличения отдельного участка спектра. На вкладке "Информация" (рис.3) в виде таблицы располагается информация, полученная после полной обработки масс-спектра, со значениями следующих параметров для каждого пика:

- M – массовое число ионов вещества;
- Масса – точная масса ионов вещества (положение центра пика), в а.е.м.;
- Ошибка – среднеквадратичное отклонение массы.
- Me – дефект массы (разность массы и массового числа), в mmu (тысячная доля а.е.м.).
- DMe – разность между текущим и предыдущим Me, в mmu.
- R – безразмерный параметр, определяющий форму пика.
- dR – среднеквадратичное отклонение параметра R.
- Высота – высота пика, отсчитанная от уровня фона, ион/с.
- Std Err – стандартное отклонение МНК для данного пика, деленное на высоту пика (относительное стандартное отклонение).
- Iter – количество итераций, осуществленных МНК для данного пика.
- Sum – сумма квадратов отклонений точек от аппроксимирующей кривой.
- Kol – количество точек для данного пика.

Программа позволяет сохранить данную таблицу в HTML-формате.

M	Масса	Ошибка	Me	DMe	R	dR	Высота	Std Err	Itel	Sum	Kal
295	295.015220	0.002076	15.23		2590	116.8	23626	0.08	16	422083726.04	110
296	296.309204	0.001381	309.20	294.0	2184	56.1	2912	0.05	17	2145370.21	101
297	297.264089	0.006198	264.09	-45.1	1546	131.5	694	0.16	100	1240587.59	101
298	298.144850	0.008869	144.85	-119.2	732	73.0	392	0.19	46	505191.44	95
299	299.257017	0.002593	257.02	112.2	2352	119.6	1502	0.09	18	2094683.65	108
300	300.266731	0.002660	266.73	9.7	2361	124.1	1321	0.10	20	1514519.47	98
301	301.278022	0.001497	278.02	11.3	2319	66.8	2710	0.06	17	2441617.43	107
302	302.287795	0.001581	287.80	9.8	2303	69.7	2108	0.06	17	1516679.10	101
303	303.299284	0.001809	299.28	11.5	2296	79.0	2010	0.07	18	1776751.62	102
304	304.308581	0.001662	308.58	9.3	2297	72.6	1968	0.06	18	1395608.57	99
305	305.319892	0.002252	319.89	11.3	2374	104.2	1810	0.08	20	2209773.02	103
306	306.329934	0.001582	329.93	10.0	2325	70.1	2445	0.06	18	2027769.21	101
307	307.336605	0.002396	336.61	6.7	2343	107.5	1668	0.09	21	2007307.90	101
308	308.349550	0.001159	349.55	12.9	2297	49.8	3415	0.04	17	2228795.14	104
309	309.356513	0.001216	356.51	7.0	2251	50.3	3286	0.05	17	2121920.70	100
310	310.357880	0.002438	357.88	1.4	2179	94.6	1261	0.08	22	1040113.56	98
311	311.046773	0.002732	46.77	-311.1	1820	89.3	498	0.08	21	83796.81	51
311	311.314338	0.002504	314.34	267.6	1672	68.8	398	0.07	8	30367.88	48
312	312.154434	0.010417	154.43	-159.9	660	78.8	307	0.22	59	440503.14	97
313	313.301827	0.004366	301.83	147.4	2552	227.7	805	0.15	26	1313819.45	94
314	314.065347	0.000683	65.35	-236.5	2432	32.1	36580	0.03	15	100295094.54	106
315	315.321110	0.002004	321.11	255.8	2327	86.5	1542	0.07	20	1218677.24	102
316	316.329288	0.001701	329.29	8.2	2254	69.1	1686	0.06	18	994568.86	98
317	317.338462	0.002311	338.46	9.2	2456	110.0	1374	0.08	21	1280032.80	101
318	318.346206	0.001939	346.21	7.7	2346	84.5	1626	0.07	19	1221755.23	99

Рис.3. Вкладка "Информация"

На вкладке "Тестирование" размещена общая информация по обработке: количество обработанных пиков в данном спектре, количество проблемных пиков, которые не были обработаны вследствие того, что во время работы МНК параметры вышли за допустимый диапазон и приняли нереальные значения (это может произойти, если высота пика близка к амплитуде флуктуаций, или если пик является частью двойного

пика), время вычислений, потраченное на обработку всего спектра, параметры проблемных пиков, параметры всех зон пиков.

На вкладке "Калибровочные массы" (рис.4) пользователь может задать калибровочные (реперные) массы. Калибровка спектра делается для того, чтобы учесть изменение магнитного поля, которое имело место при регистрации спектра.

Все массы

Калибровочные массы

Ввод вручную

Стандартные массы

Создать кривую

Очистить список

296.34462, C 21 H 44, молек
310.36018, C 22 H 46, молек
324.37584, C 23 H 48, молек
338.3915, C 24 H 50, молек
352.40716, C 25 H 52, молек
366.42282, C 26 H 54, молек
380.43848, C 27 H 56, молек
394.45414, C 28 H 58, молек
408.4698, C 29 H 60, молек
422.48546, C 30 H 62, молек
436.50112, C 31 H 64, молек
450.51678, C 32 H 66, молек
464.53244, C 33 H 68, молек
478.5481, C 34 H 70, молек
492.56376, C 35 H 72, молек
506.57942, C 36 H 74, молек
520.59508, C 37 H 76, молек

Рис.4. Вкладка "Калибровочные массы"

Те массы, которые пользователь выберет в качестве калибровочных из списка "Все массы", заносятся в левую часть списка "Калибровочные массы" по возрастанию величины масс.

В группе "Ввод вручную" пользователь набирает или выбирает из списка стандартных истинное, по его мнению, значение для массы, отмеченной в списке калибровочных масс, и вводит его в последний столбец списка "Калибровоч-

ные массы". После того, как пользователь ввел калибровочные массы, он должен нажать на кнопку "Создать кривую", при этом будет создана калибровочная кривая, которая отображается на вкладке "Калибровочная кривая".

На вкладке "Теория СН" (рис.5) отображаются дефекты масс для углеводородов в виде теоретической и практической кривой.

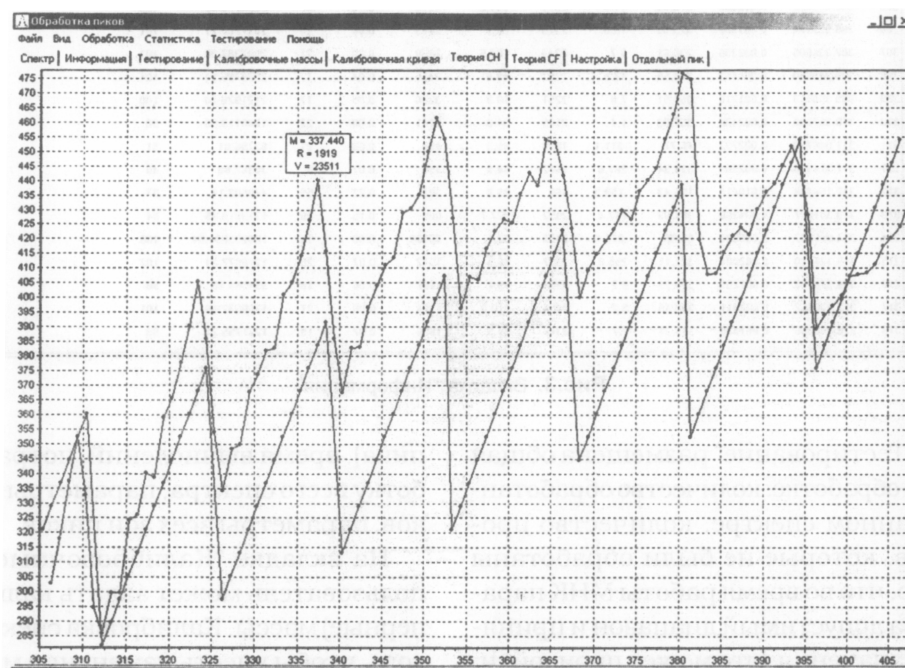


Рис.5. Вкладка "Теория СН"

Теоретическая кривая имеет форму "пилы", верхние точки которой соответствуют молекулярным углеводородам (C_nH_{2n+2}). Между молекулярными углеводородами идут углеводороды с соответствующим количеством отщепленных атомов водорода. На вкладке "Теория CF" аналогично вкладке "Теория СН" отображаются отклонения от массового числа для масс фторуглеродов.

На вкладке "Настройка" можно произвести предварительную и последнюю обработки. Для проведения предварительной обработки необходимо задать значения KolUsred и KrivoMn в соответствующих полях ввода (KolUsred – количество усреднений спектра, KrivoMn – множитель, который указывает во сколько раз одна из сторон зоны пика должна быть больше другой, чтобы идентифицировать эту зону как часть непомятого пика).

Вкладка "Отдельный пик" сделана для возможности обработки отдельного пика и получения информации о параметрах для каждой итерации МНК с целью сравнения этой информации с информацией, выдаваемой программой

TableCurve 2D Windows v.2.01 (AISN Software), которая позволяет аппроксимировать один отдельный пик функцией, задаваемой пользователем.

В программе имеется возможность сохранить изображение калибровочной кривой, графиков с вкладок "Теория СН" и "Теория CF" в файле формата BMP, возможность просмотра графиков всех стадий обработки.

На масс-спектрометре МИ-1201АГМ в режиме счета ионов был снят и обработан спектр остаточных углеводородов в диапазоне 300-400 а.е.м. Установлено, что фон спектра имеет периодичность 14 а.е.м. Следует заметить, что "пила" углеводородов тоже имеет периодичность 14 (масса CH_2 равна 14 а.е.м.) и высоты пиков имеют такую же периодичность. В программе в качестве аппроксимирующей функции для пика спектра, как изложено выше, взята функция Гаусса. При рассмотрении же фона был сделан вывод, что форма пика не является гауссовской и по краям пик спадает с меньшей крутизной (почти линейно).

В Интернете была найдена зарубежная программа IsoPro 3.0 (размещается на странице

<http://members.aol.com/msmssoft>) фирмы MS/MS Software. Она служит для отображения и расчёта интенсивностей пиков веществ с тяжёлыми изотопами, имеющих заданный атомный состав. Расчёты показали, что у пика Лоренца фон по краям уменьшается медленно. Таким образом, возможно, что в спектре углеводородов пики являются пиками Лоренца. Формулы пиков даны в [2, с. 82]. В этих формулах учтён угловой разброс ионов. Дополнительное влияние на форму пика оказывает объёмный заряд [2, с. 46].

Если в спектре присутствуют углеводороды и фторуглероды, то из-за близости масс их пики частично накладываются друг на друга. Было проверено, что при обработке, которая реализована в программе, рассчитанное положение центра каждого пика в двойном пике отличается от положения центра каждого пика при обработке двойного пика по формуле двух пиков на 10-20 мтм. Это вызвано тем, что каждый пик своим спадом влияет на другой, из-за чего на результирующем спектре их центры расположены ближе друг к другу, чем на самом деле. Поэтому корректнее обрабатывать двойной пик по формуле двух пиков (обозначения аналогичны приведенной выше функции Гаусса):

$$f(M) = A + B \cdot e^{-\left(1.6651 \cdot R \cdot \left(\frac{M}{M_0} - 1\right)\right)^2} + C \cdot e^{-\left(1.6651 \cdot R \cdot \left(\frac{M}{M_0} - 1\right)\right)^2} \quad (2)$$

Определение наличия одного пика или двух и обработка спектра по соответствующим форму-

лам является перспективой дальнейших разработок.

Дополнительно был обработан спектр в диапазоне масс 12-37 а.е.м. и для каждого пика был определен соответствующий ему состав ионов (в том числе были обнаружены ионы H_2O , LiO). Также был обработан спектр углеводородов и спектр, содержащий углеводороды и фторуглероды. Распределение углеводородов хорошо совпало с теоретическим.

Распределение же фторуглеродов показало, что массы пиков больше теоретических, из чего были сделаны выводы: либо из-за обработки двойного пика не по формуле двух пиков центры пиков фторуглеродов и углеводородов сближаются, либо во фторуглеродах вместо некоторых атомов фтора присутствуют атомы водорода.

Результаты тестирования и результаты обработки масс-спектров показали, что использование программы ускоряет расшифровку масс-спектров. Перспективы продолжения данной работы заключаются в реализации возможности аппроксимации двойных пиков по формуле двух пиков, устранении ошибок, которые возникают при обработке отдельных линий и спектров с малым отношением сигнал-шум, автоматизации расшифровки спектров, рассмотрении аппроксимации пиков функцией Лоренца, учете веса для отдельных точек экспериментального пика, улучшении алгоритмов обработки, оптимизации вычислений, применении технологии нейронных сетей.

ЛИТЕРАТУРА

1. Щиголев Б.М. Математическая обработка наблюдений. М.: Наука, 1969. 344 с.
2. Сысоев А.А. Физика и техника масс-спектрометри-

ческих приборов и электромагнитных установок. М.: Энергоатомиздат, 1983. 256 с.

* * * * *

PROGRAM FOR PROCESSING OF MASS-SPECTRUM

A.A.Solomein, B.A.Kalinin, L.M.Aleksandrov

A program for processing of mass-spectrum, representation of information about parameters of peaks, for calibration of mass-scale and for representation of mass-defects of hydrocarbon and fluorine-carbons has been developed. Algorithm of processing and interface of program was viewed.